

SDS 383D Chapter 1: Preliminaries

Bayesian inference in simple conjugate families

Xizewen Han

Jan 27, 2019

(A)

Sol: For $i = 1, \dots, N$ such that $x_i \sim \text{Bernoulli}(w)$, we have

$$p(x_i|w) = w^{x_i}(1-w)^{1-x_i}$$

Therefore with prior $w \sim \text{Beta}(a, b)$, we have the posterior distribution

$$\begin{aligned} p(w|x_{1:N}) &\propto p(x_{1:N}|w) \cdot p(w) \\ &= \prod_{i=1}^N p(x_i|w) \cdot p(w) \\ &= w^{\sum x_i} (1-w)^{N-\sum x_i} \cdot \frac{1}{B(a, b)} w^{a-1} (1-w)^{b-1} \\ &\propto w^{\sum x_i + a - 1} (1-w)^{N - \sum x_i + b - 1} \end{aligned}$$

Note that the last expression is the kernel of $\text{Beta}(\sum_{i=1}^N x_i + a, N - \sum_{i=1}^N x_i + b)$ distribution, which by adding the normalizing term we'd have this as the posterior distribution. ■

(B)

Sol: For such transformation, which is one-to-one and onto, we have the inverse as

$$\begin{cases} x_1 = y_1 y_2, \\ x_2 = y_2 - y_1 y_2 \end{cases}$$

Therefore, we have the Jacobian as

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{bmatrix} = y_2 \geq 0$$

thus the joint distribution is

$$\begin{aligned}
f_{Y_1, Y_2}(y_1, y_2) &= f_{X_1, X_2}(x_1(y_1, y_2), x_2(y_1, y_2)) |J| \\
&= f_{X_1}(x_1(y_1, y_2)) f_{X_2}(x_2(y_1, y_2)) |J| \\
&= \frac{1}{\Gamma(a_1)} (y_1 y_2)^{a_1-1} e^{-y_1 y_2} \cdot \frac{1}{\Gamma(a_2)} (y_2 - y_1 y_2)^{a_2-1} e^{-y_2 + y_1 y_2} \cdot y_2 \\
&= \underbrace{\frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} (1 - y_1)^{a_2-1}}_{g(y_1), \text{ kernel of Beta}(a_1, a_2)} \underbrace{y_2^{a_1+a_2-1} e^{-y_2}}_{h(y_2), \text{ kernel of Gamma}(a_1+a_2, 1)}
\end{aligned}$$

Since the joint distribution of Y_1 and Y_2 can be factored into two functions, $g(y_1)$ and $h(y_2)$, by Lemma 4.2.7 of Casella and Berger's Statistical Inference, as well as the form of each function noted above in the last expression, we conclude that Y_1 and Y_2 are independent random variables such that

$$Y_1 \sim \text{Beta}(a_1, a_2)$$

and

$$Y_2 \sim \text{Gamma}(a_1 + a_2, 1)$$

Based on such conclusion, we propose the following method to simulate a $\text{Beta}(a_1, a_2)$ random variable:

- 1st step: simulate from two Gamma random variables: $X_1 \sim \text{Gamma}(a_1, 1)$ and $X_2 \sim \text{Gamma}(a_2, 1)$
- 2nd step: calculate $Y_1 = \frac{X_1}{X_1 + X_2}$

Such $Y_1 \sim \text{Beta}(a_1, a_2)$. ■

(C)

Sol: With $X_i \sim \mathcal{N}(\theta, \sigma^2)$ for $i = 1, \dots, N$, in which θ is unknown and σ^2 is known, and $\theta \sim \mathcal{N}(m, v)$, we have the posterior distribution

$$\begin{aligned}
p(\theta | x_{1:N}) &\propto p(\theta) \cdot p(x_{1:N} | \theta) \\
&= \frac{1}{\sqrt{2\pi}v} e^{-\frac{(\theta-m)^2}{2v}} \cdot \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sum (x_i - \theta)^2}{2\sigma^2}} \\
&\propto \exp\left(-\frac{(nv + \sigma^2)\theta^2 - 2(m\sigma^2 + v(\sum x_i))}{2\sigma^2 v}\right)
\end{aligned}$$

By noting the kernel, we recognize the posterior distribution as

$$\mathcal{N}\left(\frac{m\sigma^2 + v(\sum_{i=1}^N x_i)}{\sigma^2 + nv}, \frac{\sigma^2 v}{\sigma^2 + nv}\right)$$

The variance term can be re-written as

$$\frac{\sigma^2 v}{\sigma^2 + nv} = \frac{1}{\frac{1}{v} + \frac{n}{\sigma^2}}$$

Note the denominator is the precision, and it's the sum of data precision and prior prevision – a good characteristic of posterior of normal-normal model. ■

(D)

Sol: With $X_i \sim \mathcal{N}(\theta, 1/w)$ for $i = 1, \dots, N$, in which θ is known and precision w is unknown, and $w \sim \text{Gamma}(a, b)$, we have the posterior distribution

$$\begin{aligned} p(w|x_{1:N}) &\propto p(w) \cdot \prod_{i=1}^N p(x_i|\theta, w) \\ &\propto w^{a-1} e^{-\frac{1}{b}w} w^{\frac{N}{2}} e^{-\frac{\sum (x_i - \theta)^2}{2} w} \\ &= w^{\frac{N}{2} + a - 1} e^{-\left(\frac{1}{b} + \frac{\sum (x_i - \theta)^2}{2}\right) w} \end{aligned}$$

By recognizing the kernel, we have the posterior distribution

$$w \sim \text{Gamma}\left(\frac{N}{2} + a, \frac{1}{b} + \frac{\sum (x_i - \theta)^2}{2}\right)$$

Since w is a re-parameterization of σ^2 , we use change of variables directly on posterior pdf to derive the posterior of σ^2 . We have the transformation $g(w) = \frac{1}{\sigma^2}$, which is one-to-one and onto, and

$$\frac{d}{d\sigma^2} g^{-1}(\sigma^2) = \frac{d}{d\sigma^2} (\sigma^2)^{-1} = -(\sigma^2)^{-2}$$

Therefore, for a general case $w \sim \text{Gamma}(a, b)$, we have

$$p(\sigma^2) = p(g^{-1}(\sigma^2)) \left| \frac{d}{d\sigma^2} g^{-1}(\sigma^2) \right| = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a-1)} e^{-\frac{b}{\sigma^2}} (\sigma^2)^{-2} = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}$$

i.e. $\sigma^2 \sim IG(a, b)$. Thus for our specific case, the posterior distribution of σ^2 is

$$\sigma^2 \sim IG\left(\frac{N}{2} + a, \frac{1}{b} + \frac{\sum (x_i - \theta)^2}{2}\right)$$

■

(E)

Sol: With known idiosyncratic variances for normal likelihood, the posterior density for unknown common mean θ is

$$\begin{aligned} p(\theta|x_{1:N}) &\propto \mathbb{P}(\theta) \cdot \prod_{i=1}^N p(x_i|\theta) \\ &= \frac{1}{\sqrt{2\pi v}} e^{-\frac{(\theta - m)^2}{2v}} \frac{1}{\sqrt{2\pi \sigma_i}} e^{-\frac{1}{2} \sum \frac{(x_i - \theta)^2}{\sigma_i^2}} \\ &\propto e^{-\frac{1}{2} \left(\left(\frac{1}{v} + \sum \frac{1}{\sigma_i^2} \right) \theta^2 - 2 \left(\frac{m}{v} + \sum \frac{x_i}{\sigma_i^2} \right) \theta \right)} \end{aligned}$$

We can thus directly write out the posterior distribution for θ as

$$\mathcal{N}\left(\frac{\frac{1}{v}m + \sum_{i=1}^N \frac{1}{\sigma_i^2} x_i}{\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2}}, \frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2}\right)$$

notice that the mean term is expressed in the form as a weighted average of the observations x_i 's and the prior mean m . ■

(F)

Sol: Using the parameterization from the question sheet, we compute the marginal distribution of x as

$$\begin{aligned} p(x) &= \int_w p(x|w)p(w)dw \\ &= \int_w \left(\frac{w}{2\pi}\right)^{1/2} e^{-\frac{w}{2}(x-m)^2} \frac{(b/2)^{a/2}}{\Gamma(a/2)} w^{a/2-1} e^{-\frac{b}{2}w} \\ &= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \cdot \int_0^\infty \underbrace{w^{\frac{a+1}{2}-1} e^{-\left(\frac{b}{2} + \frac{(x-m)^2}{2}\right)w}}_{\text{kernel of Gamma}(\frac{a+1}{2}, \frac{b+(x-m)^2}{2})} dw \\ &= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \frac{\Gamma(\frac{a+1}{2})}{\left(\frac{b+(x-m)^2}{2}\right)^{\frac{a+1}{2}}} \cdot 1 \\ &= \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})} \frac{1}{\sqrt{\pi b}} \left(1 + \frac{(x-m)^2}{b}\right)^{-\frac{a+1}{2}} \\ &= \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})} \frac{1}{\sqrt{a\pi}(b/a)^{1/2}} \left(\frac{a + \left(\frac{x-m}{(b/a)^{1/2}}\right)^2}{a}\right)^{-\frac{a+1}{2}} \end{aligned}$$

Note that the last expression is exactly the pdf of t location-scale distribution with $d = a$ degrees of freedom, center m , and scale parameter $(b/a)^{1/2}$. ■